Animal number versus maximum tumour volume: an example of reduction and refinement trade-off in the 3Rs

Dominique-Laurent Couturier[1] and Henri G. M. J. Bertrand[2]*
[1]Cancer Research UK - Cambridge Institute, University of Cambridge Li Ka Shing Centre, United Kingdom
[2]University Biomedical Services, University of Cambridge, United Kingdom

***e-mail:** dominique-laurent.couturier@cruk.cam.ac.uk

To the Editor — The 3Rs - Replacement, Refinement, Reduction - are well established guiding principles, outlined by Russel and Burch in 1959 [1], applied in animal research to improve animal welfare when no alternative to animal use is available. Extensions of the 3Rs to ensure scientific validity have been suggested. Examples are the additional 3Rs - Robustness, Registration and Reporting - of Strech and Dirnagl [2], and the 3Vs - construct, internal and external Validity - of Wuerbel [3]. The internal consistency and compatibility of these principles of animal welfare and scientific validity have been recently evaluated by Eggel and Wuerbel [4] who note that reduction and refinement are the only two components amongst the 3Vs and 3Rs where conceptual conflicts exist.

The purposes of reduction and refinement are respectively to use the minimum number of animals necessary in order to achieve scientific endpoints as well as to minimise the negative experience animals will undergo during research projects (https://www.nc3rs.org.uk/the-3rs). These aims are interdependent and can oppose each other: Refining a humane endpoint may require increasing the number of animals needed to reach the scientific endpoints. Decreasing the number of animals may require lowering the experimental animal's quality of life by pushing the humane endpoint to achieve the same level of precision. Known examples of conflicts between these two principles are the use of longitudinal (instead of cross-sectional) studies, the re-use of animals, and the amount of information collected for each animal [4, 5, 6].

Thus, the trade-offs between the reduction and refinement principles can be at the centre of discussion during the ethical review of research projects. While reviewing project licence applications related to cancer studies in our Institute, we noted that the peer-review published guidelines used by applicants to define their humane endpoint sometimes caused an increase in the number of animals needed to achieve the desired statistical power, i.e., to detect the effect of interest with a given level of certainty (typically 80 or 90%). We would like to highlight the trade-off that exists between planned maximum authorised subcutaneous tumour volume per animal and animal numbers when considering longitudinal tumour growth analyses.

The left plot of Figure 1 shows the simple scenario we considered. The average tumour volumes for treated (violet) and untreated (pink) mice are compared. We assumed that tumour growth is a power function of time from enrolment and treatment, and that the treatment would induce, on average, a 25% tumour volume reduction after 6 weeks. Parameters of this example were derived from the study of Mohamed et al [7] considering mice subcutaneously transplanted with MCF7 breast cancer cells and twice a week calliper-based tumour volume estimations assuming spheroid tumour shapes [8,9].

The right plot of Figure 1 shows combinations of maximum tumour volume limit in mm3 (x-axis) and total sample size (y-axis), leading to a power of 80% (pink) and 90% (violet) to detect

the difference in growth of interest. We can clearly see the interdependence between refinement and reduction: the lower the tumour volume limit, the larger the sample size needed to achieve the targeted statistical power, and vice versa. In the case we considered, the number of animals required to achieve an 80% or 90% power sharply decreases when the maximum tumour volume rise from 500 to 750 mm3. However, the decrease in sample size induced by considering tumour volume larger than 1500 mm3 appears more limited. More specifically, a sample size of n=32 mice would be required if the maximum tumour size was set to 1000 mm3 to achieve a 90% power. This roughly corresponds to tumours of 11 x 13 mm diameter, recommended as the maximum size for subcutaneous tumours in mice by the National Cancer Research Institute Guidelines [10], and is often set as a tumour burden endpoint. However, the guidelines allow for a higher maximum, up to 15 mm mean diameter, for therapeutic studies only. We find that if the maximum tumour volume is increased to 1500 mm3, roughly corresponding to tumours of 13 x 15 mm diameter, a sample size of n = 22 mice would be needed. Thus, for 90% power, an increase of 50% in the maximum tolerated tumour volume would lead to a 31% decrease in animal number. A similar trade off occurs when considering an 80% power.

Such trade-offs between refinement and reduction are expected to occur in other areas and to be specific to the scenario of interest. They will change depending on factors like, for example, the effect size of interest, the targeted power level, the within-mouse dependence levels, the number of weekly measures per animal and the chosen statistical technique. In the case that initiated this reflection, the Institute's ethic committee assessed that the tumour burden would be 'similar' for tumour volumes ranging from 1000 to 1500 mm^3 as long as the tumour did not incur any complications. Therefore, a large maximum tumour size limit allowed one to decrease the animal number required in order to achieve the targeted power without increasing the animal discomfort.

Due to the strong relationship between the principles of reduction and refinement, we argue that they must be assessed as a function of one another for the determination of appropriate humane and scientific endpoints during the ethical review process of research proposals. A comprehensive understanding of tumour development behaviour and its consequence on animal physiology and welfare are essential to conduct this process.

Acknowledgments

We thank Dr. Davina Honess at CRUK – Cambridge Institute for her valuable suggestions which helped improve the letter as well as the Carroll group at CRUK – Cambridge Institute for providing the raw data of [7] used to derive the scenario considered in our power analysis.

Competing interests

The authors declare no competing interests.

[1] Russell, W.M.S. and Burch, R.L. (1959), *The Principles of Humane Experimental Technique*, Methuen, London.
[2] Strech D. and Dirnagl U. (2019). *BMJ Open Science*, **3**:e000035.
[3] Wuerbel H. (2017). *Lab Anim. (NY)*, **46**, 164–166.
[4] Eggel, M. And Wuerbel H. (2021). *Laboratory Animals*, 55(3) 233–243.
[5] Boo, M.J., Rennie, A.E., Buchanan-Smith, H.M. et al. (2005). *Animal Welfare*, **4**, 327–332.
[6] Olsson, I.A.S, Franco, N.H., Weary, D.M., et al. (2011). *ALTEX proceedings of the 8th world congress on alternatives and animal use in the life sciences, Montreal 2011*, Johns Hopkins University Press, Baltimore, 333–336.
[7] Mohammed, H. et al. (2015). *Nature*, 523(7560), 313-317. doi:10.1038/nature14583
[8] Euhus, D.M., Hudd, C., LaRegina, M.C. and Johnson, F.E. (1986). *Journal of Surgical Oncology*, **31**, 229–234. doi: 10.1002/jso.2930310402.
[9] Tomayko, M.M. and Reynolds, C.P. (1989). *Cancer Chemotherapy and Pharmacology*, **24**, 148–154.
[10] Workman P., Aboagye E.O., et al. (2010). *British Journal of Cancer*, **102**, 1555–1577.

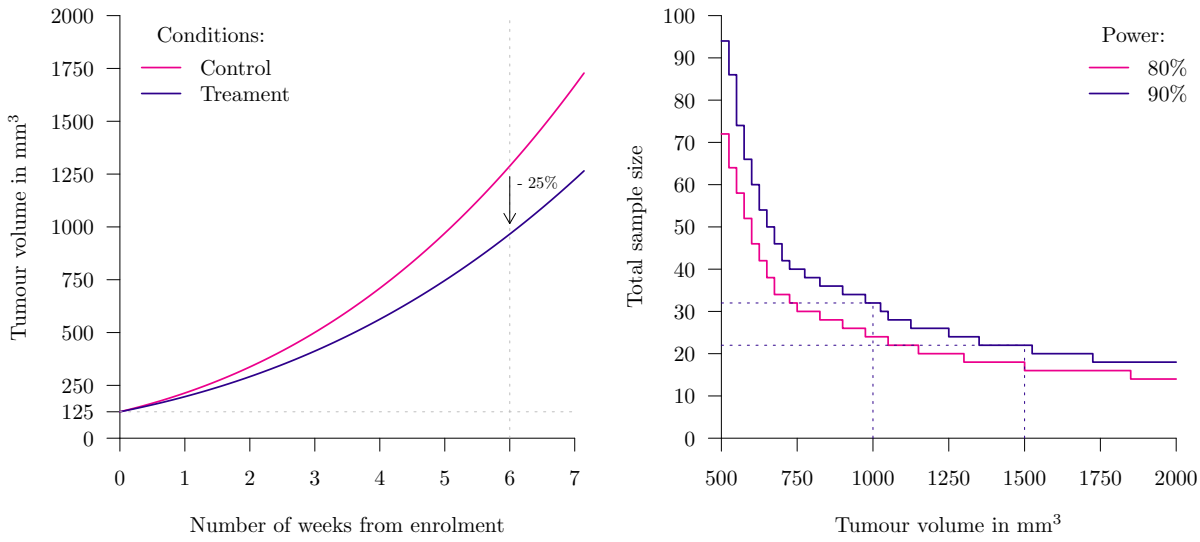

Fig. 1| Power analysis effect size and results.
Left panel: Relationship between the average tumour volume (y-axis) and the time from enrolment (x-axis) per treatment group (coloured lines) assumed in the power analysis. Right panel: Total sample size (y-axis) required to detect the average difference in growth displayed in the left panel with a probability of 0.8 (pink) and 0.9 (violet) as a function of the maximum tumour volume limit in mm3 (x-axis).